



Time-frequency analysis with temporal and spectral resolution as the human auditory system

Agerkvist, Finn T.

Published in:

Proceedings of the IEEE-SP International Symposium of Time-Frequency and Time-Scale Analysis

Link to article, DOI:

[10.1109/TFTSA.1992.274127](https://doi.org/10.1109/TFTSA.1992.274127)

Publication date:

1992

Document Version

Publisher's PDF, also known as Version of record

[Link back to DTU Orbit](#)

Citation (APA):

Agerkvist, F. T. (1992). Time-frequency analysis with temporal and spectral resolution as the human auditory system. In *Proceedings of the IEEE-SP International Symposium of Time-Frequency and Time-Scale Analysis* (pp. 425-428). IEEE. <https://doi.org/10.1109/TFTSA.1992.274127>

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

TIME-FREQUENCY ANALYSIS WITH TEMPORAL AND SPECTRAL RESOLUTION AS THE HUMAN AUDITORY SYSTEM

Finn T. Agerkvist

The Acoustics Laboratory, Technical University of Denmark, Building 352
DK-2800 Lyngby, Denmark

ABSTRACT

The human perception of sound is an obvious area for application of simultaneous time-frequency analysis since the ear is selective in both domains. A perfect reconstruction filter bank with bandwidths approximating the critical bands is presented. The orthogonality of the filters makes it possible to examine the masking effect with realistic signals.

The tree structure of the filter bank makes it difficult to get well attenuated stopbands, but the use of filters of different length solves this problem.

1. INTRODUCTION

Extensive work on the understanding of the perception of sound by man has shown that the human auditory system is selective in both time and frequency. However, most work in the field has concentrated on examining either the temporal or the spectral resolution only, this has been partly due to the necessity of reducing the complexity of the investigations and partly due to the lack of a established tool for simultaneous time and frequency analysis. In the last years simultaneous time and frequency analysis has received much attention and a promising application of this analysis seems to be in the field of understanding and modelling the human ear.

2. DESIGN OF A FILTER BANK WITH RESOLUTION PROPERTIES LIKE THOSE OF THE HUMAN EAR

The filter bank described in this paper is based on the tree-structured use of a two-channel perfect reconstructing quadrature mirror

filter bank (PR-QMF).

2.1 The Two-Channel PR-QMF Bank

The design of a two-channel PR-QMF bank is well known [1], it is based on a half-band low pass analysis FIR filter, $h_0(n)$, of odd order, $N-1$, the z -transform of which is $H_0(z)$. The corresponding high pass analysis filter is chosen as $h_1(n) = (-1)^n h_0(N-1-n)$, $H_1(z) = -H_0(-z^{-1})z^{N-1}$. Since the filters reduce the bandwidth to half the original, the sample rate may be compressed by a factor of two. In figure 1 the entire filter bank can be seen including the expanders and the reconstructing filters $G_0(z) = -H_1(-z)$ and $G_1(z) = -H_0(-z)$, these two filters serve the purpose of cancelling the aliasing caused by reducing the sampling rate.

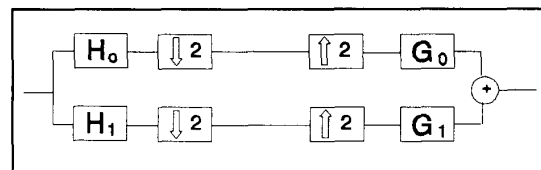


Figure 1. The two-channel quadrature mirror filter bank.

The filters have an important property namely orthogonality in time:

$$\sum_n h(n) h(n-2m) = \begin{cases} 0, & m \neq 0 \\ 1, & m = 0 \end{cases} \quad (1)$$

This has the consequence that each output sample from the compressor represents a unique part of the energy in the signal. In applications related to the human hearing this a very useful property since one may easily modify the signal at any point (strictly speaking an area around the point) in the time-frequency plane and then reconstruct the signal.

2.2 The PR-QMF Bank as a Time-Frequency Tool

The relationship between wavelet analysis and the two-channel perfect reconstruction quadrature mirror filter bank has been pointed out in recent papers [2,3]. The coefficients that satisfy the scale-function dilation-equation

$$\phi(x) = \sum_k 2 c_k \phi(2x-k) \leftrightarrow \quad (2)$$

$$\Phi(f) = \prod_{j=1}^{\infty} C(f/2^j)$$

where

$$C(f) = \sum_k c_k e^{jkf} \quad (2a)$$

and define a wavelet:

$$\psi(x) = \sum_k 2 (-1)^k c_k \phi(2x-k) \leftrightarrow \quad (3)$$

$$\Psi(f) = \Phi(f/2) \tilde{C}(f/2)$$

where

$$\tilde{C}(f) = \sum_k (-1)^k c_k e^{jkf} \quad (3a)$$

are also the coefficients of a PR-QMF low pass filter. If a PR-QMF bank with these wavelet

coefficients is used in a tree-structured filter bank design, as outlined in figure 2, and only the branches containing the lower half of the frequency range are subject to further filtering, then this filter bank will be the digital equivalent of wavelet analysis, the similarity can be seen from equation 2 and 3: The scale function is the infinite product of scaled versions of a low pass filter, and the wavelet is a scaled version of a high pass filtered scale function. It is possible to draw the output anywhere in the tree instead of only at the high pass filtered branches: The output from the column of filters to the right on figure 2 gives a time-frequency analysis with the three properties: i) constant bandwidth, ii) orthogonality in time, iii) an acceptable simultaneous time-frequency resolution. (This combination is not possible with the short-time Fourier transform.)

2.3 The Masking Effects

If a time-frequency analysis of an acoustic signal is to show the signal as it is perceived by the ear, postprocessing becomes necessary because of the masking effects. Psychoacoustic experiments [4] have demonstrated the frequency masking effect: A pure tone (or a narrow band of noise) may mask a pure tone (or a noise band) of low level. The ability to mask is constant within the so-called critical bandwidth around the (center)frequency of the masking component, outside this band the ability to mask decreases. The critical bands lies close to that of a one-third octave filter at high frequencies, at lower frequencies, i.e. below 800 Hz, the critical bandwidth is approximately equal to 100 Hz.

A high level signal component may also mask a short duration signal appearing either before or after the masking component, this is called pre/post-masking.

According to the models for masking effects the time-frequency spectrum can be modified so that the inaudible energy is removed from the signal, in this way it is possible to test the masking model with more realistic signals than pure tones and random noise. It is only possible to modify the time-frequency spectrum and still be able to reconstruct the signal because the filters have the perfect reconstruction property.

With an initial sampling frequency of 25.6 kHz the bandwidths shown in table 1 may be obtained with a tree structured filter bank

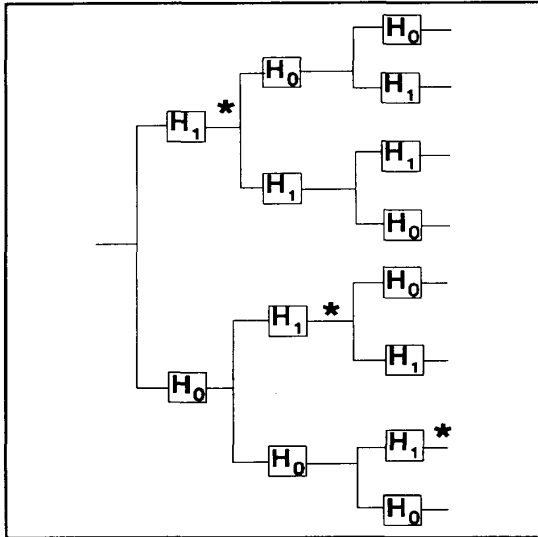


Figure 2. The analysis part of a tree-structured filter design with compressors omitted. The branches marked with an * contain the output from a wavelet analysis.

design. For comparison, table 1 also shows the critical bands. The filter bank has been implemented on a IBM-PC (Turbo-Pascal).

Because of the tree structure there is a potential problem with overlap between the high

Filter Bank		Critical Band	
f_c	Δf	f_c	Δf
50	100	50	100
150	100	150	100
250	100	250	100
350	100	350	100
450	100	450	110
550	100	570	120
650	100	700	140
750	100	840	150
900	200	1000	160
1100	200	1170	190
1300	200	1370	210
1500	200	1600	240
1800	400	1850	280
2200	400	2150	320
2600	400	2500	380
3000	400	2900	450
3600	800	3400	550
4400	800	4000	700
5200	800	4800	900
6000	800	5800	1100
7200	1600	7000	1300
8800	1600	8500	1800
10400	1600	10500	2500
12000	1600	13500	3500

Table 1. Bandwidths of proposed filter bank compared to critical bands.

and low pass filters: it is important to use filters with steep slopes outside the pass band. It is particularly important that the first PR-QMF filter banks, i.e. those placed near the root of the tree-structure, use very sharp filters since the frequency response of these filter covers a large frequency range. Unfortunately it seems that the maximum obtainable simultaneous time and frequency resolution of a wavelet (or a PR-QMF bank) decreases slightly as the number of coefficients is increased [5]. Figure 3 shows a frequency sweep analysed with the frequency resolution shown in table 1 and with a 22 tap filter optimized for maximum simultaneous time-frequency resolution [5]; as seen on the figure the overlap causes some strange looking mirror images, and as expected these images are largest around 6400 kHz. In figure 4 the total frequency response of the filter covering 4000 Hz to 4800 Hz is shown. The overall response is the product of one lowpass filter followed by three high pass filters, all four filters uses the 22 tap filter.

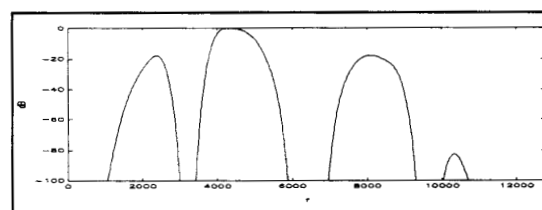


Figure 4. The total filter response for band 16: With a 22 tap filter in all 4 filters.

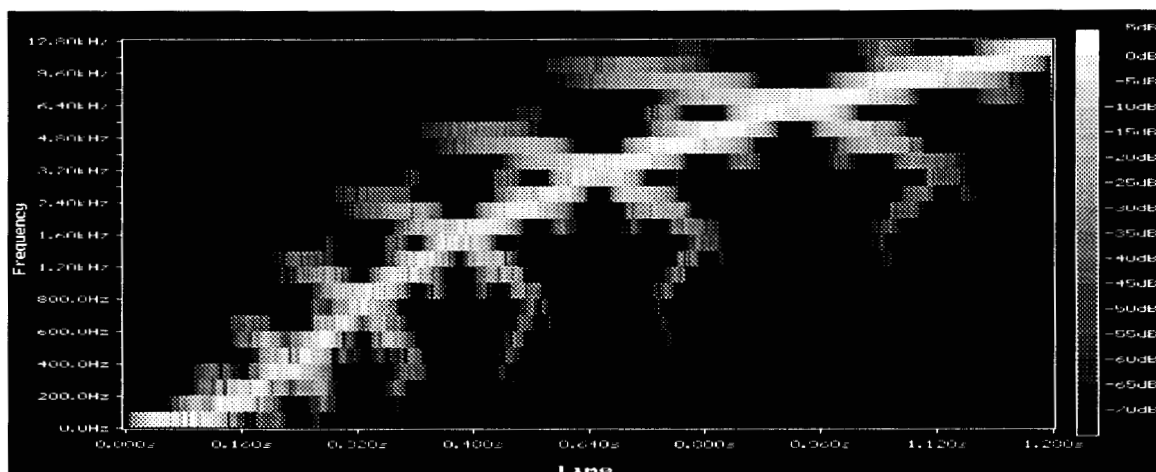


Figure 3. The Time-Frequency spectrum of a sinus sweep. Note the strange looking mirror images appearing around the crossover frequencies caused by the finite slopes of the filters.

The figure shows several only slightly attenuated ripples and these ripples explain the mirror images seen on figure 3.

3. DISCUSSION

The necessity of the simultaneous time and frequency analysis in modelling the masking phenomenon may be illustrated by the following example: A signal consisting of two pure tones with approximately the same frequency and the same high level and a low level pure tone with a quite different frequency is analysed with the ordinary Fourier transform using a long time window. According to such analysis the tone of low level will be masked, but since the two high level tones have almost the same frequency their sum will be a beating tone, and during the cancelling period the low level tone can be heard.

The at present available models for the masking effect are relative simple and one may question the validity of these models since: i) the focus has not been on both time and frequency, ii) the psychoacoustic experiments have been carried out with a limited number of combinations of signals, iii) the signals used in the experiments have had little similarity to music or speech. A filter bank as described here may be used to obtain more detailed information on the masking effect.

As mentioned in section 2.3 the simultaneous time-frequency resolution decreases as the length of the filters increases. This is not a serious problem; since the final resolution is determined by the filters that are furthest from the root one may choose to use the long filters near the root only, where a steep slope is needed. This of course reduces the similarity to the wavelet analysis. Figure 5 shows the frequency response of the filter shown in figure 4 when a

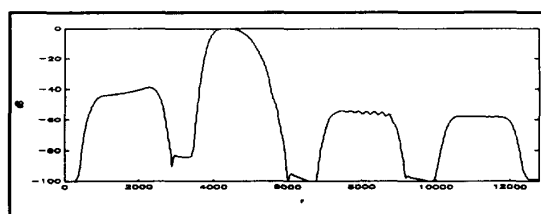


Figure 4. The total filter response for band 16: With a 70 tap filter in the first 2 filters and with a 22 tap filter in the last 2 filters.

70 tap filter is used the two first times. It can be seen that the ripples that caused the mirror images on figure 3 are reduced without changing the passband.

4. APPLICATIONS

A filter bank as described here combined with an accurate model for the masking phenomena has important and promising possibilities of application: it may replace time consuming and expensive listening tests in testing non-linear and/or time-invariant equipment such as: 1) the DCC-tape recorder with a data reduction system, 2) hearing aids with compressors, 3) telephone systems with adaptive filters and speech detectors.

5. CONCLUSIONS

A filter bank with resolution properties like those of the human ear has been designed and implemented using perfect reconstruction quadrature mirror filter banks in a tree structure. The frequency resolution of the filter bank is approximately that of the critical bands. The orthogonality of the filters used makes it possible to simulate the masking phenomenon and to test the model of masking with realistic signals. With a valid model for the masking phenomenon the filter bank may replace listening test.

References:

- [1] Smith, M. J. T. and Barnwell, T. P. : Exact Reconstruction Techniques for Tree-Structured Subband Coders. *IEEE Trans. on Acoustics, Speech and Signal Processing* June 1986, 434-441.
- [2] Rioul, O. and Vetterli, M. : Wavelets and Signal Processing. *IEEE Signal Processing Magazine* Oct. 1991, 14-38.
- [3] Daubechies, I.: *Orthonormal Bases of Compactly Supported Wavelets*. *Comm. on Pure and Applied Mathematics* Vol. XLI, 909-996.
- [4] Zwicker, E.: *Psykoakustik*, 1982. Springer Verlag, Berlin.
- [5] Doritze, C. and Villemoes, L. F. : Optimizing Time-Frequency Resolution of Orthonormal Wavelets. *Proceedings of IEEE ICASSP 1991*, 2029-2032.